A Novel and Automatic Circular Contour Approximation Based Breast Tumor Classification Algorithm

M.S.Abdaheer[#1], Ekram Khan[*2]

Aligarh Muslim University,
Aligarh, India
sss.viji@gmail.com,sss_viji@yahoo.com

***Abstract*— In this paper a novel and automatic method for classification of breast malignancy in digital mammogram images is proposed. It is based on two features named as Normalized mean radial distance as F_1 and number of crossing point between fitting circle and tumor for the circle as F_2. The roughness of the breast tumor out-layer is measured in terms of normalized mean radial distance. The fitting circle is obtained by considering the centroid of tumor as its centre and arithmetic mean of maximum and minimum radial distances of contour points from the centroid, as its radius. The similarity between tumor and its circle is measured in terms of number of crossing points between tumor contour and circle for the tumor contour. The simulation results show that for a set of 150 tumor contours, the proposed method gives 92.67% accuracy for F1 and 94.667% accuracy for F2. The performance obtained in terms of the receiver operating characteristic (ROC) parameters like accuracy (A_c), sensitivity (S_e), specificity (S_p), and positive (PPV) and negative predictive values (NPV) for F1 and F2 as(0.92667,0.9241,0.9296,0.9359,0.9167),(0.94667,0.9873,0.9014,0.9176,0.9846) respectively.**

***Keywords*—**Breast tumor, Benign, Malignant, Circular approximation method, Normalized mean radial distance, Number of crossing points

I. INTRODUCTION

Breast cancer is the common diseases among women who have crossed 50 and above. It pushes the cervical cancer in to second spot. The death mortality rate is increased every year as per the survey report displayed by the American cancer society [1]. These surveys are counting only on metros but not count in to the rural areas. Early detection and characterization of breast cancer can apparently improves the situation and possibly reduce the chances of false biopsies. Many methods like Digital Mammography, MRI and Ultrasound scanning are commonly used for early detection of breast cancers [3]. However, the failure to detect any abnormal lesion at an early stage may lead to disastrous consequences. The mammographic images show signs of obstruction and many direct and indirect radiographic signs due to space occupying lesions in the tissue region of breast. The interpretation of poorly illuminated mammography is very difficult and physicians with different levels of experience can come up with different results for the same breast mammography images. To minimize the operator related deficiency many image processing based algorithms are developed for assisting the radiologist to analyze the breast mammography. Through this image processing based technique, features for classification are extracted from the input image [4]. With the help of classifier, tumors are classified as benign or malignant based on their features values and its decision threshold.

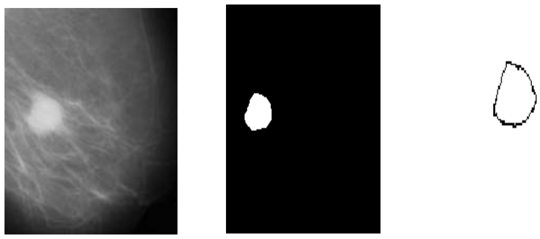
S. Baeg et al classified the breast abnormalities based on textural features named as denseness and architectural distortion and obtained satisfactory performance [5]. Rangayyan et al. used acutance as feature for classification, achieving a classification rate of 92.6% for manually traced 54 tumor contours [4]. El-Faramawy used various shape features like compactness, Fourier descriptor, moments and chord-length statics for classifications. However, combination of all these shape features could achieve only 76% classification efficiency[6]. Menut et al. classified the breast masses with the help of parabolic modeling method with the accuracy of 76% [7]. Arun Thaitaikumar et al. reduced the 56.3% of biopsies by using axial shear estrography as a classifier for classification of breast tumors [8] . S.C.Yanga et al. detected and classified the breast masses with the help of five texture and four shape features as inputs of Parabolic neural network classifier and achieved 84.15% classification rate with overall area under ROC of 0.93 [9].

H. W. Lee et al. analyzed the infiltrative nature of breast masses by octave energy derived from reversible round off non recursive 1-D Discrete periodic wavelet transform. A test dataset of breast sonograms with the lesion contour delineated by an experienced physician and three datasets of breast sonograms with the lesion contour delineated by a Java-based image processing program, ImageJ, are built for feature efficacy evaluation. They achieved 95.1% and 84.4% of accuracy for manual and Image J generated data base by combining octave energy feature with morphometric feature [10]. Bi Lio et al developed local texture based fully automated breast tumor classification system with the accuracy of 93.75% [11]. T. M. Nugayan et al. achieved 94.6% of classification accuracy with the help of fractal dimensions based classifications method. Moreover, mammographic mass boundaries are usually blurry, making it difficult for automated detection schemes to detect precisely the mass boundaries. For these reasons, features quantifying the textural information contained by masses are needed to perform classification independent of shape [12].

In the present study, we focus on the development of texture measures for the classification of mammographic masses as benign or malignant. The rest of the paper is organized as follows. Section 2 describes about materials and methods. Section 3 describes about result and discussion. Section 4 will finally conclude the paper.

II. MATERIALS AND METHODS

The necessary data set for testing and implementation of our algorithm is taken from MIAS data base [2], R.M.Rangayyan data base [4] as well as some of the contour from our local hand drawn data base. The given input mammogram image is enhanced with the help of morphological opening filter with disk structuring element of order 5. The enhanced image is segmented through intensity histogram based optimum threshold technique [13]. The ROI (region in which tumor is present) is automatically traced out from segmented foreground image. The edges of the automatically traced ROI are calculated through canny edge detector . The artifacts or high frequency noises which are present over on the extracted edges are removed through pre processing technique. In this work, morphological opening filter with disk structuring element of order 3 is used as a pre processing filter to improve the visual impactness of extracted contour image (extracted edges) [14]. The improvement of visual impactness is measured in terms of peek signal to noise ratio. The pre processed contour act as an input image of the proposed breast tumor classification algorithm as shown in Fig.1.



Input Image Identified ROI Extracted Contour

Input Image Identified ROI Extracted Contour

Fig. 1. Contour Extraction from Digital Mammogram Images

The main objective of this algorithm is to draw a circle for the given input tumor. By keeping centroid of the tumor as the center of the circle, circle is fitted on the given tumor. The radius (R) of the circle is obtained as an arithmetic mean of minimum and maximum radial distance which is measured from centroid of the tumor to coordinate of the contour. The normalized mean radial distance value depends up on the roughness of the counter. By keeping centroid of the tumor as a center of the contour, the circle is fitted on the tumor. The fitted circle may cross and merge with different points of the given tumor. The number of crossed point between fitted circle and tumor for the circle is depends on the roughness of the out layer (contour) of the tumor. The procedure to draw the circle and determination of its parameters are discussed in the following sub sections.

*A. **Centroid Calculation***

For a given contour, its centroid is calculated by approximating contour into polygon with N vertices. The coordinates (x_i, y_i) of each pixel of the contour (C) act as vertices of image [15]. If vertices of N points are known then the coordinate of centroid is calculated as follows;

$$x_c = \frac{\sum_{i=1}^{N} mixi}{\sum_{i=1}^{N} mi}; \qquad y_c = \frac{\sum_{i=1}^{N} miyi}{\sum_{i=1}^{N} mi} \tag{1}$$

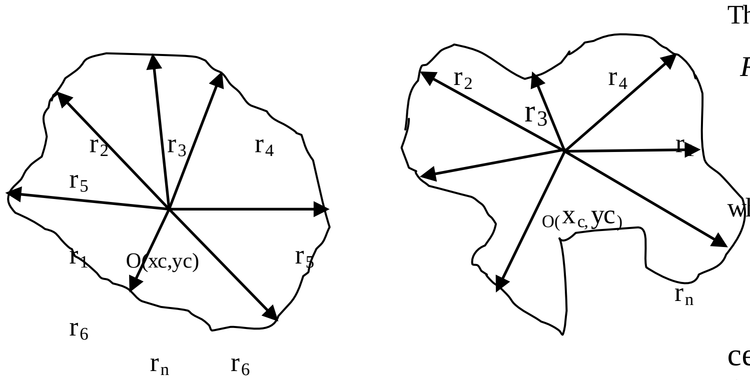
where

N is the total number of pixels in a given contour image m_i is the weighting factor (Number of white pixels in row containing the vertex) x_c and y_c are the coordinates of the centroid.

B. Converting Radial Distance as the Radius of the circle

Once the centroid (O) of the tumor is obtained, it is used as the centre of the best fitting circle. The next step is to find the radius of the best fitting circle, which is the arithmetic mean of maximum and minimum radial distance between the centroid and the contour as shown in the Fig.2. The radial distance, r(x, y), between i^{th} contour point (x_i, y_i) and centroid (x_c, y_c) can be measured as follows:

$$r(x,y) = \sqrt{(x_c - x_i)^2 + (y_c - y_i)^2} \quad \forall (x_i, y_i) \in C \quad i = 1,2..N-1, N \tag{2}$$



(a) (b)

Fig.2 Radial distances measurement of (a). Benign tumor and (b). Malignant tumor

As evident fromFig.2, that measured radial distance value depends upon the roughness of tumor contour. The smooth and round shaped contours of the benign tumors likely to have larger value of normalized mean radial distance, whereas rough structured and complex natured malignant tumor likely to have less value of normalized radial distance. Therefore normalized mean radial distance used as a feature for classification of breast tumors.

The normalized mean radial distance (N_{mr}) of the given tumor is defined as follows.

$$N_{mr} = \frac{1}{N} \sum_{x=1}^{n} \sum_{y=1}^{n} r(x,y) \tag{3}$$

In the proposed algorithm, the normalized mean radial distance defined in Eqn. (3) is used as feature for the classification of tumors in either benign class or malignant class, by comparing the feature value $F_1 = N_{mr}$ with a pre-determined threshold (T_1) (to be discussed in the next subsection). As stated earlier that benign tumors have smooth, round or oval shaped contour, and thereby larger value of F_1, whereas the malignant tumors having rough and ill structured contour surface will have small value of F_1. Therefore, for a given threshold (T_1) and a contour with feature value F_1, the classifier works as follows;

$$\begin{aligned} &if\ F_1 < T_1 \Rightarrow \quad malignant\ \ tumor; \\ &else \qquad \Rightarrow \quad benign \quad tumor \end{aligned} \tag{4}$$

The arithmetic mean of minimum and maximum radial distance defined in Eqn (2) used as a radius of the best fitting circle.

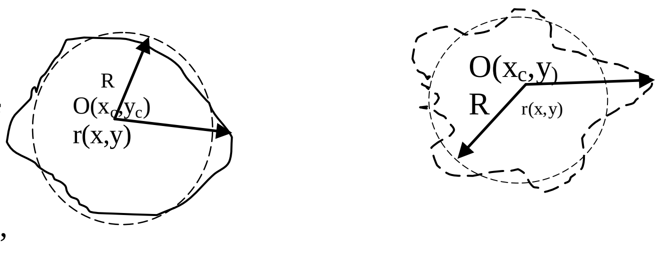
Then the radius of best fitting circle is

$$R = \frac{r_{\max} + r_{\min}}{2} \tag{5}$$

where

$$r_{\max} = \max_{\forall (x,y) \in C} r(x,y)$$
$$r_{\min} = \min_{\forall (x,y) \in C} r(x,y)$$

By keeping centroid of a given tumor as a center of circle and radius calculated from Eqn. (5), the circles for benign and malignant tumors are shown in Fig. 2.



(a) (b)

Fig 3.The Best Fitting Circles for contours of (a) Benign Tumor and (b) Malignant Tumor

C .Calculating Number of Crossing Points (N_c)

As evident from Fig. 3 that the overlapping area between a given contour and corresponding fitting curve depends on the roughness of the contour. The smoother contours of benign tumors are likely to have less number of crossing points with the fitting circle, whereas the contours with rough outlier of malignant tumors are likely to have more number of crossing points with the circle. Therefore, the number of crossing points between circle and tumor for the circle is used as a feature for classification of breast tumors.

For exactly count the number of crossed points $(N_1, N_2, N_3......N_{n-1}, N_n)$, number of merged points (N_m) are calculated and subtracted from number crossed and merged points (N_{cm}) .

Number of Crossed Points (N_c)

$$= \quad (N_1, N_2, N_3......N_{n-1}, N_n) \quad ;$$

$$1 \le n \le N-1, \quad \forall N \in C(x,y) \cap Cir(x,y) \qquad (6)$$

$$N_c = N_{cm} - N_m$$

where

N_{cm} = Number of crossed and merged points

N_m = Number of merged points

Numbers of crossed and merged points are calculated as follows

1.The given input image I(x, y) is scanned from top to bottom and left to right.
2.If pixel p of given input image I(x, y) is one, then it will pass in to next pixel.
3.If pixel p of given input image I(x, y) and its four neighborhood pixel value is equal to zero then it will check its radial distance

$$if \;\; p(x,y) = 0 \quad then$$

$$p(x,y) = \begin{Bmatrix} p(x+1,y) = 0;\, p(x-1,y) = 0 \\ p(x,y-1) = 0;\, p(x,y+1) = 0 \end{Bmatrix} \quad \forall (x,y) \in C \qquad (7)$$

4. The radial distance of corresponding pixel is not equal to radius of fitted circle I_1 (x, y), then it will be counted as a number merged points.

$$if \quad p_1(x,y) \cup p(x,y)$$

$$count = 0$$

$$then$$

$$\begin{Bmatrix} r(x+1, y, r(x-1, y) \\ r(x, y-1), r(x, y+1) \end{Bmatrix} \le R \qquad (8)$$

$$\begin{Bmatrix} r(x+1, y), r(x-1, y) \\ r(x, y-1), r(x, y+1) \end{Bmatrix} > R \qquad (9)$$

$$Count = Count + 1 = N_m \qquad (10)$$

where

P_1 is the pixel of the fitted circle
P is the pixel of given input breast tumor contour
x & y is the coordinates of the pixel.

Total number of pixel whose values is zero also counted as total number of merged and crossed points (N_{cm}).

5. Total number of crossed point is calculated by subtracting total number of merged points from total number of points.

$$N_c = N_{cm} - N_m \qquad (11)$$

In the proposed algorithm, number of crossed points (N_c) defined in Eqn (11) is used as feature for the classification of tumors in either benign class or malignant class, by comparing the feature value $F_2 = N_c$ with a pre-determined threshold (T2) (to be discussed in the next subsection). As stated earlier that benign tumors have smooth contour, and thereby smaller value of F_2, whereas the malignant tumors having rough contour surface will have large value of F2. Therefore, for a given threshold (T) and a contour with feature value F, the classifier works as follows;

$$if\ F_2 \geq T_2 \Rightarrow \ malignant\ tumor;$$
$$else \quad \Rightarrow \ benign \ tumor \qquad (12)$$

D. Decision Threshold

It is evident from Eqn (4) and Eqn.(12) that for classification, a threshold value (T) is to be calculated for classification. For a given training set of extracted contours of tumors (with known class and feature value F), the threshold value (T) is calculated for Feature F_1 and F_2 as follows.

The decision threshold for F_1 is calculated as arithmetic mean of maximum value of F_1 in malignant class and minimum value of F_1 in benign class, which is as follows

$$T = \left(\frac{F_{1M,\max} + F_{1B,\min}}{2} \right) \qquad (13)$$

where,

$F_{M,\ max}$ is the maximum feature value of malignant class with in training data set.

$F_{B,\ min}$ is the minimum feature value of benign class with in training data set

The decision level for F_2 is calculated as arithmetic mean of maximum value of F_2 in benign class of training set and minimum value of F in the malignant class, which is as follows

$$T = \left(\frac{F_{2B,\max} + F_{2M,\min}}{2} \right) \qquad (14)$$

where,

$F_{B,\ max}$ is the maximum feature value of benign class with in training data set.

$F_{M,\ min}$ is the minimum feature value of malignant class with in training data set

Based on this decision level, tumors (training and validation sets) are classified according to Eqn (4). and Eqn.(12). The performance of the classifier is evaluated through Receiver Operating Characteristics (ROC). In ROC, if malignant tumor is classified as malignant, it is called as a true positive case (TP). If malignant is classified as benign or benign is classified as malignant then it is called as a false negative (FN) and false positive (FP) respectively. If benign class is predicted as a benign then it's called as true negative (TN). Accuracy (Acc), hit rate (TPR), false alarm rate (expense) (FPR), specificity (Sp), positive predictive value (PPV) and negative predictive value (NPV) of classifiers are calculated for measuring the efficiency of classifiers.

$$Acc\ (A_c) = \frac{TP + TN}{TP + FP + TN + FN} \qquad (15)$$

$$senitivity\ (se) = TPR \ = \frac{TP}{TP + FN} \qquad (16)$$

$$specificity\ (sp) = 1 - FPR \ = \frac{TN}{FP + TN} \qquad (17)$$

$$PPV = \frac{TP}{TP + FP} \qquad (18)$$

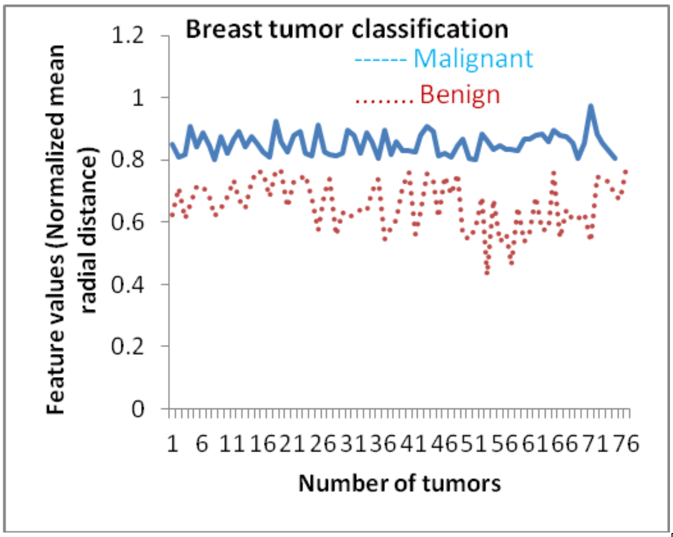
$$NPV = \frac{T\ N}{FN + TN} \qquad (19)$$

III. RESULT AND DISCUSSION

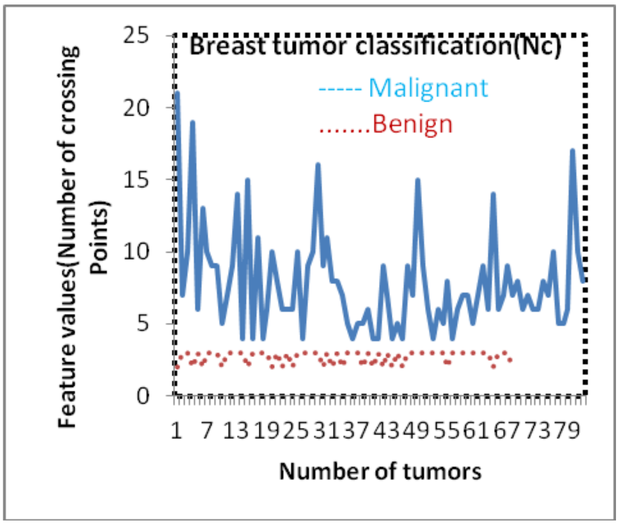
The performance of the proposed algorithm is evaluated for a total of 150 tumors (90 as a training set and 60 for validation and verification). Out of these 150 tumors, there were 71 benign tumors and 79 malignant tumors. Features (normalized mean radial distance (N_{rm}), Number of crossing points (N_c)) for classification is extracted through our proposed algorithm for all 150 breast tumor contours. Decision threshold defined in Eqn.(13) (for F1) and Eqn.(14) (for F_2) is calculated for 90 contours of training set (with known feature value and class) through stastical based optimum threshold technique and is found to be 0.8 ,03 for F_1 and F_2 respectively.

Fig. 4 shows the feature values (or defined in Eqn.(3) & Eqn(11)) for tumors classified as benign and malignant. It can be observed that there exists a clear cut demarcation between the feature values of two classes of tumors at decision threshold values. Table 1 and Table 2 shows the performance of the proposed algorithm (based on F1 and F2 respectively) for all 150 tumors in terms of classification accuracy defined in Eqn. (15). It can be observed from Table 1 that the proposed method results into misclassification of only 5 out of 71 benign tumors and only 6 out of 79 malignant tumors, giving the overall accuracy of 92.66%. Similarly based on second feature (number of crossing points (F_2)), the proposed algorithm performance is summarized in Table 2 It can be observed from Table 2 that the proposed algorithm misclassified only 5 out of 71 benign tumors and 6 out of 79 malignant tumors ,

giving the overall accuracy of 94.667%. Table 3 and Fig. 5 compare the performance of the proposed method with other existing methods in terms of overall classification accuracy defined in Eqn.(15) . It is evident from the table and figure that the accuracy of the proposed method is approximately 10-19% better than its contemporary methods based on F_1 and F_2.



(a)



(b)

Figure 4. Classified tumors with its Feature (Variance) Value (a) Based on F1; (b) Based on F2

The performance of the proposed method is also evaluated in terms of ROC parameters such as A_c, s_e, s_p, PPV, NPV defined in Eqn (15)-(19). The ROC parameters for the proposed method and other similar methods are summarized in Table 5. It can be observed from the table that the proposed method outperforms than other methods consistently for all ROC parameters, which show the ability of our algorithm to efficiently classify the breast malignancy.

TABLE 1. Breast tumor classification Based on Normalized Mean radial distance (F1)

Input Data's	Case	Based on Normalized mean radial distance		
		Classified as		% Accuracy
		Benign	Malignant	
Benign	71	66	05	92.95%
Malignant	79	06	73	98.73%
Total	150	72	78	92.66%

TABLE 2 Breast tumor classification Based on Number of crossing points(N_c)

Input Data's	Case	Based on Normalized mean radial distance		
		Classified as		% Accuracy
		Benign	Malignant	
Benign	71	66	05	92.95
Malignant	79	03	76	96.20
Total	150	69	81	94.667

TABLE 3. Performance Comparison of proposed method with the other existing methods

S. No	Name Author'(s) and their Proposed Method		Percentage of Accuracy
1	Oliver Menut et.al/ Parabolic Modelling Method[3]		76%
2	N.R. Mudigonda et.al /Iterative Boundary Segmentation algorithms[11]		81%
3	W.H.lee et al /Wavelet Based Breast tumor Classification Method[6]		84%
4	Proposed Method	Based on F1	92.667%
		Based on F2	94.667%

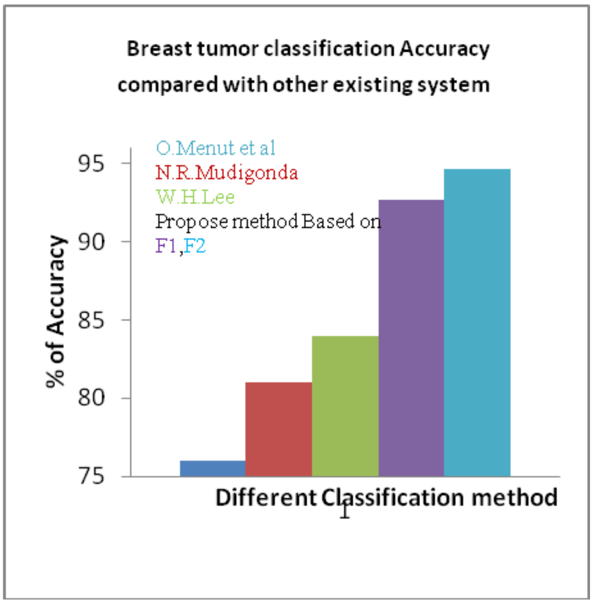


Figure 5. Comparison with other existing System (in terms of Accuracy)

Table 5. Comparison of ROC parameters with other existing system

Measured ROC parameters	Wavelet Based Method	Local texture Method	Our Proposed Method	
			Based on F_1	Based on F_2
A_c	0.84	0.9375	0.92667	0.94667
S_e	0.933	0.95	0.9241	0.9873
S_p	0.795	0.9231	0.9296	0.9014
PPV	0.714	0.9344	0.9359	0.9176
NPV	0.956	0.9412	0.9167	0.9846

CONCLUSION

In this paper we proposed a simple novel and automatic algorithm for breast tumor classification. The classification is based on two features which can be used to measure the roughness of the outlier of the breast tumor. The simulation results showed that the proposed efficiently classifies a wide range of tumors and outperforms other methods in terms of ROC parameters. Another salient feature of this method is that it is simpler compared to other method as only centroid and radius of the circle are need to be calculated.

Acknowledgment

The authors wish to acknowledge the help provided by Prof. R. M. Rangayan, Department of Electrical and Computer Engineering, Department of radiology, university of Calgary, Canada, for providing necessary database to test our algorithms.

Reference

[1]. American cancer society home page. Available: www.cancer.org

[2]. Digital Mammogram Image Analysis homepage .Availabale:http://www.mammoimage.org/databases/

[3]. Sue Moss, "Screening for Breast cancer in India- Is it an appropriate strategy?" journal of national cancer institute, vol.100, pp.1270-1271, sep 2008.

[4]. R. M. Rangayyan , N.M. El-Faramawy, J. E. Leo Desautels, O. A. Alim, "Measures of acutance and shape for classification of Breast tumors", IEEE transaction on medical imaging, vol.16, Dec.1997.

[5]. S. Baeg and N. Kehtarnavaz, "Classification of breast mass abnormalities using denseness and architectural distortion", Electronic letter on computer vision and Image analysis, vol.1, pp.1-20, June 2002.

[6]. N. M. El-Faramaway , R. M. Rangayyan, J. Desautels, O. Alim , "Shape factors for analysis of breast tumors in Mammograms," in proc. CCECE-96, 1996, pp. 355-358.

[7]. O. Menut, R. M. Rangayyan, J. Desautels, "Classification of breast tumors via Parabolic modelling of their contours," in.proc: PACRIM-97, 1997, pp.1002-1005.

[8]. A. Thitaikumar, L. M. Mobbs, C M Kramer-chant, B S Garra, J Ophirl , "Breast tumor classification using axial shear strain elastrography," IOP publication on Physics in Medicine and Biology , vol.53, pp.4809-4823, Sep.2008.

[9]. S. C. Yang, C. M. Wang, Y. N. Chung , G C Hsu, S K Lee, P C Chung, C I Chang, "A computer aided system for mass detection and classification in digitized mammograms, J. Bio Medical Engineering Applications, Basic &Communication vol. 17, pp. 215-228, June. 2005.

[10]. H. W. Lee, B. D. Liu, K. C. Hung, S. F. Lei , P. C. Wang, T. L. Yang, "Breast tumor Classifications of Ultrasound Images using Wavelet Based Channel Energy and Images J, IEEE journals on selected topics in Signal Processing, vol. 3, pp. 81-93, Mar. 2009.

[11]. B. Lio, H. D. Cheng, J. Hung, J. Liu, "Fully automatic and segmentation –robust classification of breast tumors based on Local texture analysis," in proc. JCIS-08, 2008, pp. 1-7.

[12]. T. M. Nguyan, R. M. Rangayan, "Shape analysis of breast masses in mammograms via the Fractal dimensions," in proc.EMBC'05, 2005, pp. 3210-3213.

[13]. Frank Chung-Hoon Rhee, Y.S.Shin, "A Fast Numerical Method for Finding the Optimal Threshold for Image Segmentation," in proce. of Fuzz '03, 2003, pp. 984-989.

[14]. R. A. Peters, "A new algorithm for image noise reduction using mathematical morphology," IEEE transactions on Image processing, vol. 4, pp. 554-568, May 1995.

[15]. Lee. N. J., "Computer-Aided Diagnostic Systems for Digital Mammograms, Master of science dissertation, Louisiana State University, Louisiana, Dec. 2006.